THE MATHEMATICS TEACHER

EDITED BY

W. H. METZLER

ASSOCIATED WITH

EUGENE R. SMITH   HARRY D. GAYLORD

GEO. GAILEY CHAMBERS   WILLIAM E. BRECKENRIDGE

VOLUME X   JUNE, 1918   NUMBER 4

SOME APPLICATIONS OF MATHEMATICS TO EDUCATIONAL STATISTICS.*

By GEO. GAILEY CHAMBERS.

At the end of the first semester of the academic year 1916–1917, there were 22,912 grades assigned by the instructing staff in the University of Pennsylvania to the students in the undergraduate departments, that is, the students in the College, the Towne Scientific School, the Wharton School and the School of Education. There were five grades used, namely, D (Distinguished), G (Good), P (Passed), N (Not passed) and F (Failed). The significance of these grades, considered as percentages, varied in the different departments of instruction. There was no attempt at uniformity in this respect, nor was there any effort at uniformity in the percentage of D grades, G grades, etc.

The percentages of the different grades were as follows:

D's	14
G's	35
P's	38
N's	9
F's	3

There seems to be justification for assuming that each of these grades is the result of a measurement of some common trait, to which the term general ability might wisely be applied.

* Read before the meeting of the Association of Teachers of Mathematics of the Middle States and Maryland, December 1, 1917.

It is true that each measurement was a very complicated process beginning at the first session of the class at the opening of the semester, and continuing until the term examination was given at the end of the semester. It is also true that the process varied very greatly in different cases. For example, in some cases, the students whose progress was measured were seniors, and in others they were freshmen. Also, for example, in some cases the subject-matter of a technical laboratory course in electrical engineering was used in the measuring process; while in other cases, the subject-matter of a course in business law was used.

Nevertheless I believe we may assume that these 22,912 measurements were distributed according to the familiar normal probability curve as expressed by the equation

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

σ being the standard deviation for the whole set of measurements, x being the value of an individual measurement, and y the corresponding frequency of that particular measurement.

An essential condition for the application of the theory of probability to any set of statistics is that there shall be operating "numerous independent causes." Surely that is the case here.

Moreover, each group of students taken together in one application of the measuring process is approximately homogeneous in its stage of educational progress; that is, each section or class brought together for instruction is approximately homogeneous, at least in comparison with the crowd that gathers on the street when the fire bell rings, for example. Also the subject-matter and methods of presentation are adjusted to the stage of educational progress attained by the group.

Furthermore, speaking generally, we know of no cause that would tend to skew the distribution of the measurements in one of these instruction groups from the normal distribution. It is possible, however, that the standard deviation would vary with different groups, whereas, the assumption that the normal law of errors holds for the distribution of the whole set of 22,912 measurements implies the assumption that the standard deviation is the same for all of the instruction groups.

Therefore the results obtained from the use of the probability curve in such a case must be considered as very rough approximations. Nevertheless, I believe the considerations just given justify its use. The results are only slightly less reliable than are the grades themselves, and every college professor feels that much reliance can be placed on the grades he gives, and deans and executive committees feel sufficient reliance upon them to justify their dropping students from college, thus stopping their educational career along one line at least.

The area under the frequency curve will then represent the 22,912 grades. A table like that on page 219 in Thorndike's "Introduction to the Theory of Mental and Social Measurements" shows that the ordinates corresponding to $x = -1.83\,\sigma$, $-1.15\,\sigma$, $0.01\,\sigma$, and $1.06\,\sigma$ will divide the area into portions corresponding respectively to the percentages of F's, N's, P's, G's and D's.

Also the ordinates corresponding to $x = -2.12\,\sigma$, $-1.41\,\sigma$, $-0.49\,\sigma$, $0.47\,\sigma$ and $1.46\,\sigma$ will divide the areas for the F's, N's, P's, G's and D's into equal parts respectively.

By taking a new origin $3\,\sigma$ to the left, a new unit equal to one tenth of σ and dropping the right-hand figure so as to be within the limits of accuracy of our data, we have the following as average values of the grades: $F = 9$, $N = 16$, $P = 25$, $G = 35$ and $D = 45$.

This computation was carried out to obtain the best values of the grades to use in ranking the undergraduate students, either all together or by groups. I applied these results first to the group of freshmen who entered directly from their secondary schools; that is, with less than a year of time between the completion of their school work and the beginning of their university work. There were 680 such freshmen.

In determining the values of the grades I decided to use the percentages of grades assigned to the whole undergraduate student body, rather than the percentages assigned to any smaller group. The corresponding percentages assigned to the group of freshmen just mentioned were as follows:

D's	12
G's	34
P's	39
N's	10
F's	5

If these percentages had been used to determine the values of the grades and if the origin had been properly chosen, we would have had the same average values for the grades. I believe now that it would have been better to have used only the grades of the freshmen, on the assumption that the common trait whose measure is represented by the grades would be approximately the same in amount for the freshmen, but would be higher for upper classmen. Fortunately, as it happened, the use made of these results is valid whichever set of grades is considered as the basis for determining the values of the respective grades.

Having these values of the grades and having the percentages of the different grades obtained by the individual students, we computed a rank number for each freshman. For example, a certain student's record showed 33 per cent. G's, 56 per cent. P's and 11 per cent. N's. His rank number was 274. The rank numbers would run from 90 to 450. They were computed by multiplying the percentages of the respective grades by the values of the corresponding grades, dropping the right-hand figure, and adding the results.

The whole set of students was then arranged according to the rank numbers, and separated into three groups: first, those who were in the upper quarter of the arrangement; second, those in the lower quarter; and third, those between the first two groups. Each student was then given a rank letter, U, M or L, according to whether he was in the upper, middle or lower group.

It would evidently be unwise to use the rank numbers as an accurate determination of individual ranks, because the data upon which the computation is based would not justify such precise results. I believe, however, that the rank letters, L, M and U, are sufficiently reliable for many purposes.

If now we assume that the 680 students whose ranking we have determined are distributed according to the normal law of frequency, then a table of values of the probability integral shows that the average values of L, M and U may be taken as 9, 20 and 31 respectively, with a fair degree of approximation.

Now take any group of students out of the whole set; for example, those who offered Latin for entrance; determine the percentages of those students that ranked in the L, M and U

groups respectively, and multiply those percentages by the corresponding values of those ranks and add the products, first dropping the right-hand figure to keep within the range of accuracy of the data. If the sum is more than 200 the group stands higher than a normal group, but if the sum is less than 200 the group stands lower than normal. The following tables give illustrations of such group standings.

RANKING OF GROUPS OF STUDENTS WHO ENTERED THE UNIVERSITY OF PENNSYLVANIA IN SEPTEMBER, 1916, DIRECTLY FROM SECONDARY SCHOOLS, BASED ON GRADES ASSIGNED IN ALL SUBJECTS AT END OF FIRST SEMESTER.

1. *Grouped According to Amount of Foreign Language Credited on Entrance.*

Amt. Lang.		No. with Rank			Total.	Amt. Lang.	No. with Rank			Total.	Amt. Lang.	No. with Rank			Total.
		L.	M.	U.			L.	M.	U.			L.	M.	U.	
2 units	No.	25	41	13	79	5 units	30	41	16	87	8 units	1	4	4	9
	Per cent.	32	52	16			35	47	18			11	45	44	
	Pt.	29	104	50	183		31	94	56	181		10	90	136	236
3 units	No.	25	48	14	87	6 units	27	64	46	137	9 units	2	5	8	15
	Per cent.	29	55	16			20	47	33			13	34	53	
	Pt.	26	110	50	186		18	94	102	214		12	68	164	244
4 units	No.	42	82	26	150	7 units	13	58	43	114	11 units	0	1	1	2
	Per cent.	28	55	17			11	51	38			0	50	50	
	Pt.	25	110	53	188		10	102	118	230		0	100	155	255

2. *Grouped According to the Number of Foreign Languages Credited on Entrance.*

Number of Languages.		No. with Rank			Total.
		L.	M.	U.	
1.............	No.	63	123	36	222
	Per cent.	28	56	16	
	Pt.	25	112	50	187
2.............	No.	94	203	117	414
	Per cent.	23	49	28	
	Pt.	21	98	87	206
3.............	No.	8	18	18	44
	Per cent.	18	41	41	
	Pt.	16	82	127	225

3. *Grouped According to the Language Credited on Entrance.*

Language.		No. with Rank L.	M.	U.	Total.	Language.	No. with Rank. L.	M.	U.	Total.
Latin	No.	110	238	130	478	German.....	122	232	127	481
	Per cent.	23	50	27			25	48	27	
	Pt.	21	100	84	205		23	96	84	203
Greek	No.	1	20	18	39	Spanish	1	7	5	13
	Per cent.	3	51	46			8	54	38	
	Pt.	3	102	143	248		7	108	118	233
French......	No.	41	86	44	171	No Latin....	55	106	41	202
	Per cent.	24	50	26			27	52	20	
	Pt.	22	100	81	203		24	104	62	190

The first number in the column headed "Total" gives the number of students entering with the corresponding offering in Foreign Language as given in the left-hand column.

The second number in the "Total" column gives the standing of the group. The larger the number the higher the standing. The lowest possible standing would be 90 and the highest 310.

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PA.